Child-Report Data and Assessment of the Social Environment in Schools

Natasha K. Bowen
University of North Carolina at Chapel Hill

Abstract

Objectives—This study tested the quality of data collected with the online ESSP for Children from a diverse sample of 1,172 third through fifth graders.

Methods—Mplus confirmatory factor analysis (CFA) procedures for ordinal and clustered data were used.

Results—Of 80 original items, 61 loaded on 13 dimensions in a first-order model that had good fit in three random subsamples. Children in grades 3 through 5 may not be reliable reporters about neighborhood adults' caring. However, 12 statistically sound and independent dimensions related to school, peers, family, and well-being were obtained.

Conclusions—The ESSP for Children provides school staff with quality data to use in conjunction with family and teacher ESSP data to guide intervention choices in schools.

School success remains elusive for many American elementary school children. According to National Assessment of Educational Progress (NEAP, National Assessment of Educational Progress, 2009a, 2009b), only 33% of all fourth graders were proficient in reading in 2009 and 38% were proficient in math. In addition to and in spite of the low overall proficiency figures, major group differences also persist almost 10 years after the No Child Left Behind legislation was passed with the goal of improving achievement and reducing discrepancies. Among African American fourth graders, only 16% and 15% were proficient in reading and math, respectively, in 2009. For Latino fourth graders and those from families with low incomes, the numbers were only slightly better: 17% proficiency rates in reading and a little over 20% in math in each group. In comparison, according to NEAP data, among fourth graders who were non-eligible for the school lunch program 45% and 54% tested proficient in reading and math, respectively; among white students, the figures were 42% and 51%.

The potential negative consequences of poor elementary school academic performance are serious and lifelong. Failure in elementary school is a strong predictor of poor school performance in subsequent years (Baydar, Brooks-Gunn, & Furstenberg, 1993; Lloyd, 1978), and poor academic performance during the secondary school years is predictive of dropout and low educational attainment. Low educational attainment, in turn, is strongly associated with future experiences of poverty. In 2008, for example, government sources

indicate that the mean income of individuals who started but did not complete high school was $24,700 (U.S. Census Bureau, 2008), over $2000 less than the eligibility threshold for children in a family of four to receive free lunches at school (U.S. Department of Agriculture, 2007). Free and reduced price lunch program eligibility is a commonly used indicator of poverty. Only families of four with an earner holding a bachelor's degree or higher could hope on average to earn enough income (about $60,900) to be well above federally defined cutoffs for this type of assistance (U.S. Census Bureau; U.S. Department of Agriculture).

In summary, poor elementary school performance is associated with failing to complete high school or completing high school without adequate preparation for success in post-secondary education, which in turn constrains one's chances of future economic security. Aside from its long-term financial consequences, poor elementary academic achievement is also associated with later conduct disorder (Williams, Ayers, Van Dorn, & Arthur, 2004), substance abuse (Jenson, 2004), and adolescent pregnancy(Franklin, Corcoran, & Harris, 2004), and other negative outcomes (Baydar et al., 1993) of concern to school social workers.

Social environmental causes of poor performance in elementary school

According to an ecological perspective of development, a child's successful school performance at any point in time is the result of myriad ongoing and past transactional processes between the child and the social environment (e.g., Richman, Bowen, & Woolley, 2004; Sameroff, 2000). Addressing poor performance or preventing declines in performance requires acknowledging the multiple influences at play in the classroom, school, peer system, family, and community. Factors that decrease their chances of future success represent risk factors; factors that make positive outcomes more likely are promotive factors (Sameroff, 2000). Peer rejection, neighborhood danger, family poverty, lack of teacher support, and friends with antisocial behavior are examples of risk factors for negative school outcomes described in the literature (Richman, Bowen, & Woolley). Family warmth, positive social experiences at school, and caring neighbors are examples of promotive factors (Richman, Bowen, & Woolley). Given their training, school social workers are appropriate leaders in efforts to assess and address social environmental risk factors, while also drawing upon student strengths.

Shortcomings of existing assessments in schools

From an ecological perspective on development, fully understanding why a student exhibits high or low school performance requires understanding his or her experiences in the social environment. Many assessments currently used by school professionals in their efforts to improve student achievement, however, focus on documenting and describing problematic functioning (see for example, measures summarized by Corcoran & Fischer, 2000; Merrell, 2003), not surveying a broad array of possible *causes* of the inadequate functioning. For examples, school psychologists may conduct extensive testing to obtain detailed information on academic performance and ability, learning and intellectual disabilities, and developmental milestones. Teachers and parents may provide information on signs of attention problems at school and home; and student support teams may collect information from checklists and observations about peer interactions and classroom behaviors. These most common targets of assessment in schools provide little or no information on potential *causes* of academic or behavioral problems that may lie in the social environment or at the interface of student and environment. They say little, therefore, about what interventions are needed to address problems. In addition, the types of assessments just described are typically labor-intensive efforts focusing on one child. They are not designed to identify common needs among students.

Researchers have found that teachers and other school staff are likely to want and value social environmental data on their students (Bowen & Powers, 2005). In addition, school-based programs endorsed as "evidence-based" invariably address social environmental factors associated with school success (Holmbeck, Greenley, & Franks, 2004; Powers, 2005). Yet many school-based assessments are not designed to determine which social environmental factors contribute to the unsatisfactory performance of individual students or groups of students.

Importance of child report data

One feature that enhances the quality of assessments is collecting data from multiple sources (Offord et al., 1996; Pepler & Craig, 1998; Silverman & Saavedra, 2004). The observed lack of close agreement among different sources about the same phenomena that has been documented in the literature (Achenbach, McConaughy, & Howell, 1987; Offord et al.) supports the importance of obtaining multiple views. Pepler and Craig outline the strengths and limitations of different sources, including children, parents, teachers, and observers, and conclude that combining information from multiple sources provides the most complete understanding. Evidence suggests that adults in a setting may be better sources than youth on youth behavior, especially externalizing behaviors (Silverman & Saavedra). In other domains, however, such as self-perceptions and experiences in social environments, data from children themselves may be the most informative (Measelle, John, Ablow, Cowan, & Cowan, 2005; Pepler & Craig, 1998; Reynolds & Kamphaus, 2003).

Quality of child report data

In creating child-report instruments, researchers must make extra efforts to promote and demonstrate the reliability and validity of the data collected. To obtain data with desirable psychometric characteristics, instruments designed for children "should attend to developmental considerations with regard to stimulus and response complexity" (Reynolds & Kamphaus, 2003, p. 374). This attention promotes what has been called the "developmental validity" of data from children (Woolley, Bowen, & Bowen, 2004). Although the reliability—which is a prerequisite of validity--of data collected from elementary school children generally tends to be lower than for adolescents and adults, evidence of adequate levels of reliability and validity have been documented for data collected from children in grades 3 through 5 (Brock, Nishida, Chiong, Grimm, & Rimm-Kaufman, 2008) and even younger (Measelle et al., 2005). Given the value of the children's perspectives on their own well-being and experiences in the social environment, social environmental assessments aimed atuncovering risk and promotive factors for school success should include children as respondents.

The Elementary School Success Profile (ESSP)

The current study reports on confirmatory factor analyses of data from the child report component of a comprehensive, multiple-source, social environmental assessment tool designed to help schools identify and address factors contributing to poor school performance among subgroups of students. The ESSP is an online assessment of social environmental factors that contribute to the well-being and academic success of 3rd, 4th, and 5th graders (Bowen, 2006). The ESSP is based on the School Success Profile (SSP; (SSP, Gary L. Bowen, Richman, & Bowen, 2002;G. L. Bowen, Rose, & Bowen, 2005), a self-report social environmental assessment for middle and high school students. The ESSP collects data from 3rd to 5th grade students, one of their caregivers, and their classroom teachers. As with the SSP, self-report data are considered paramount for gaining a full understanding of the factors affecting student achievement (Richman et al., 2004). The ESSP for Children contains scales related to child perceptions of the social environment and

their own well-being. Caregivers and teachers complement the child-report data with information on topics for which they are the most relevant sources. Parents and teachers, for example, provide information on children's home and school behavior respectively. Parents report on family characteristics and their parenting style. Teachers report on children's attendance, academic performance, and learning behavior. An overview of the domains and dimensions assessed with all three components of the ESSP is available elsewhere (N. K. Bowen, Bowen, & Woolley, 2004), as is information on the quality of ESSP data collected from parents (Wegmann, Thompson, & Bowen, in press).

Children taking the ESSP for Children report on their own psychological well-being and their experiences with adults and peers in the neighborhood, family, peer system, and school. The survey contains 83 items, 80 of which are grouped into 12 hypothesized scales. Special emphasis was placed during the iterative development of the ESSP for Children on maximizing the reliability and validity of the data it yields. The rigorous scale development process (N. K. Bowen et al., 2004), included a literature review of dimensions relevant for children in middle childhood, feedback from child development and educational experts, multiple rounds of cognitive testing with 3rd through 5th grade children (N. K. Bowen, 2008a), and formative psychometric tests of data from multiple samples. We know of no other assessment that assesses multiple aspects of the social environment, collects data from three sources, and has automatic scoring and immediately available online reports.

The current study builds on past tests of the ESSP for Children in several important ways. Data for the analysis came from a large, new sample that over-represents low-performing students and schools, which are the most common target of ESSP assessments. Scales that had been revised based on previous exploratory factor analyses (N. K. Bowen, 2006) were tested in the more rigorous confirmatory framework of structural equation modeling. Specifically, Mplus 4.2's (L. K. Muthén & Muthén, 2004) features for ordinal, non-normally distributed, and nested data, and data with missing values were used to test the fit of the 12-factor model, which was based on previous research, theory, and the SSP.

The current study was designed to provide evidence of the adequacy of the "internal structure" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) of data collected with the ESSP for Children. The ability of child items to yield data on important dimensions of the social environment and well-being that have adequate statistical qualities (e.g., independence among factors, and meaningful variance among children) is a prerequisite to another critical area of validity: "practice-related" validity (N. K. Bowen, 2008b), or validity related to the "consequences of testing" (American Educational Research Association et al., 1999). Criteria for practice-related validity include: "Scores fill a knowledge gap of users, are relevant to the outcomes of interest in the setting, and are related to the intervention strategies available in the setting. Use of scores benefits respondents" (N. K. Bowen, 2008b). By establishing the quality of the internal structure of the ESSP for Children, we also aimed to provide evidence that scores from its dimensions could be used by school professionals to guide the selection of interventions. As stated above, many evidence-based interventions for use in schools target social environmental factors. By helping school staff identify factors that threaten student success, The ESSP can help them determine which existing interventions are appropriate for students.

Methods

Sample

For the current analyses, data collected from schools in four districts taking part in separate ESSP projects during the 2008–2009 school year were combined into one sample. All

human subjects and data use procedures were approved by the Behavioral Sciences institutional review board of the University of North Carolina at Chapel Hill. The sample consisted of 1,251 third to fifth graders in 13 elementary schools in the four districts. Districts were rural, suburban, or urban. One district was relatively high performing, one was below average, and two were low-performing relative to the rest of the state. Schools participating in the separate projects were selected by district administrators or research staff due to their low performance, with one exception described below. Low performing schools are the primary source of data because the ESSP is intended to guide group-level interventions to improve school-level performance. The recruitment of students within schools differed across the projects. All third through fifth grade children were targeted for assessment in two of the districts due to ESSP project goals in those districts. In contrast, a random selection of low performing students (students performing below grade level on state standardized end-of-year tests) participated in the below-average district. In the above-average district, all fourth graders in four schools were targeted. Three of the schools in this district were low performing schools taking part in a three-year panel study; a fourth school joined the study in the second year because redistricting led to many study participants being assigned to the new school.

Combining the data from ongoing ESSP projects provided a large and diverse sample that was appropriate for the current psychometric study. Large CFA samples are desirable for statistical reasons; the quality of SEM estimates, standard errors, and fit statistics improve with larger samples. The diversity of the combined sample was desirable because it ensured that important subgroups of intended ESSP users were represented. Scale development samples should have the "qualitative" characteristics (DeVellis, 2003) of the populations they are intended for. The current study's findings are more likely to accurately reflect the true relationships between latent constructs and ESSP indicators that exist in the population served by school social workers because members of this population (e.g., African American and Hispanic/Latino students, free and reduced lunch program participants, and children performing below grade level) are well-represented in the sample (DeVellis).

Seventy-nine children (6.3% of the full sample) had parent and/or teacher data but lacked child data. Therefore, the tests of the ESSP for Children are based on data from the remaining 1,172 cases with child data. The 79 cases excluded from the analyses did not differ from the analyzed cases in terms of gender, race/ethnicity, or free/reduced lunch program participation (nominal by nominal cross-tabulation chi square tests, $p > .05$). However, the excluded cases were more likely to be below grade level in reading and less likely to be at grade level than cases included in the analyses ($p < .05$). Similar percentages of both groups were above grade level. The difference in academic performance of the included and excluded sample members reduces the generalizability of the findings of the study to the full sample.

Data Collection

Child data were collected online on school computers in the fall of 2008. Most children complete the ESSP in less than 15 minutes. An audio option is available on every question screen for children who prefer to have items read to them. Many children prefer to use headsets and the audio option, but information about how individual children completed the online survey was not recorded. A school adult supervised students completing the ESSP for Children, helping them log in to the online survey and ensuring they entered the serial number that would link their data to the appropriate teacher and parent data.

Measures

Eighty items assessing 12 dimensions of five domains were analyzed. Hypothesized dimensions were based on established dimensions from the SSP, literature on peer and family issues especially pertinent to middle childhood, and prior exploratory factor analyses (N. K. Bowen, 2006). In the Neighborhood domain, the dimension called Neighbors Who Care contained seven items assessing child perceptions that adults in the neighborhood, for example, know them by name, wave or say hello, and watch out for children. In the School domain, the Teachers Who Care dimension contained six items assessing perceptions that the teacher, for example, listens to the child, helps when the child doesn't understand something, and cares about the child's schoolwork. Also in the School domain, the Fun Place to Learn dimension contained four items assessing, for example, child perceptions that school is fun and that the child likes things studied. A third dimension in the School domain, Fun Place to be with Other Children, contained five items assessing child perceptions, for example, of having friends to talk to at school, play with, and eat lunch with. All items in neighborhood and school dimensions were measured with a four-point ordinal response set: never, sometimes, often, and always. In addition, all of the items are worded so that higher scores are more desirable.

In the Friends domain, the Friends Who Care dimension contained five items assessing for example, child perceptions, that friends help him or her, are on his or her side, and that friends and the child have fun together. The Accepted by Other Children dimension in the Friends domain contained seven negatively worded items assessing, for example, perceptions that other kids are mean to the child or tease the child, or that the child gets picked on at school. A third dimension in the Friend domain, Friends have Good Behavior, contained four positively worded and five negatively worded items about, for example, whether friends hit other kids, tell lies, help adults, and follow the rules. One dimension in the Family domain, contained eight items assessing child perceptions, for example, of how often adults at home make the child feel special and family helps when the child feels upset. All items in these dimensions were measured with a four-point ordinal response set: never, sometimes, often, and always. For reports going to users of the ESSP, negatively worded items in the Friends have Good Behavior scale, and all items on the Accepted scale are recoded so that higher scale scores are better. In CFA, absolute values of factor loadings are not affected by coding direction, so items were not recoded for the analysis.

Four dimensions related to the domain of Health and Well-being were assessed with the ESSP for Children. The dimension Good Physical Health had eight items assessing, for example, how many days the child feels sleepy or tired; has a toothache, stomachache or headache; or feels sick. A 3-point response set is used to measure physical health: "almost never," "some days," "most days." In the current analyses, higher scores on the physical health items reflect worse health. On ESSP reports, items on the Good Physical Health scale are reverse-coded so higher scale scores are better. The dimension Positive Feelings about Self assesses children's perceptions of how often they feel, for example, they are "a nice person," "good at art," "a good student." Eight items with the 4-point frequency response set described above were used to assess this dimension. The dimension Good Adjustment was assessed with six items about how often the child, for example, feels that nobody cares about him or her, nobody listens to him or her, or feels all alone. Three ordinal response options used with these items were: "no, never," "yes, sometimes," and "yes, often." In the current analyses, higher scores on items are less desirable. On ESSP reports, items on the Good Adjustment scale are reverse-coded so higher scores are better. The fourth dimension in the domain of Health and Well-being, Knows Where to Get Support, had seven items assessing perceptions of the availability of someone who listens to or encourages the child, or someone who sees things the way the child does. These items were measured with the 4-

point "never" to "always" response set described above. Higher scores the items for Positive Feelings about Self and Knows Where to Get Support are desirable.

Data Analysis

Confirmatory factor analysis (CFA) with Mplus 4.2 (L. K. Muthén & Muthén, 2004) was used to confirm the factor structure of the ESSP for Children. Because all items on the ESSP for Children are ordinal, it was not appropriate to use the default Maximum Likelihood estimator (Bollen, 1989; Flora & Curran, 2004; Jöreskog, 2005). Instead, the Weighted Least Squares estimator adjusted for means and variances (WLSMV) was used, as recommended by Flora and Curran. In addition, the lack of independence of error terms due to the clustering of students in schools was accommodated by modeling "school attended" as a cluster variable. Mplus's full information (FIML) procedure for handling missing values was also used, which allowed all cases to be included in the analysis, even if a case was missing values for one or more variables.

The WLSMV estimator generates a polychoric correlation matrix of the observed indicator variables (for more information, see Bollen, 1989 or Jöreskog, 2005). The correlation matrix is then used as the input matrix for the analysis, along with the appropriate weight matrix. Although a confirmatory approach was justifiable given extensive previous research on the ESSP and its theoretical foundations, we ultimately resorted to the "jigsaw piecewise technique" recommended by Bollen (2000, p. 78) for identifying the best construction of each factor and appropriate number of factors. The jigsaw piecewise technique allowed us to examine the quality of factors one at a time (if they had at least the four indicators necessary for model identification), and then in combination with more and more other factors. Ultimately we tested a model containing all of the dimensions assessed with the ESSP for Children.

In addition to using the best practices related to ordinal data and clustered characteristics of the data, we tested the hypothesized model with a calibration sample and then validated the final model with two validation samples. Validating CFA findings with independent samples, especially after modifications have been made, increases confidence that findings are not sample-specific. The two validation analyses were especially important given the "model generating" (Jöreskog, 1993) approach that was used in the calibration sample.

Results from Mplus analyses using WLSMV include a χ^2 statistic, the Tucker-Lewis and Comparative fit statistics (TLI, CFI), the Root Mean Square Error of Approximation (RMSEA) and the Weighted Root Mean Residual (WRMR). The null hypothesis in structural equation modeling is that the covariance/correlation matrix implied by the model is not statistically different from the analysis matrix. Therefore, researchers seek a non-significant χ^2 value. Citing Browne and Cudeck, Kline (2005) suggests that RMSEA values less than or equal to .05 indicate close fit, while values between .05 and .08 indicate "reasonable" fit. Values of .95 or higher are desirable for the TLI and CFI (Hu & Bentler, 1999). WRMR values should be below .90 (L. K. Muthén & Muthén, 2010). When the conclusions that can be drawn about fit in analyses using WLSMV are inconsistent across fit indices, L. K. Muthén and Muthén have recommended relying most heavily on the significance level of the χ^2 statistic (2010), a rigorous standard given that the statistic is larger and therefore more likely to be significant--for large sample sizes.

In models with poor fit, we examined multiple potential sources of the poor fit; low factor loadings, indicator R^2, residual correlations, and modification indices. Items that had any combination of red flags, such as, factor loadings under .40, squared multiple correlations well under .50 (we used a value of .33), multiple residual correlations under .10, and multiple modification indices above the threshold we specified (.5), were considered for

removal from factors. Internal consistency reliability coefficients obtained with SPSS (SPSS Inc., 2005) were also examined before items were removed from scales. Although we had not anticipated removing items from scales, the strong loadings and adequate reliabilities for shorter scales that were obtained in the analyses suggested that the desirable conditions of increased parsimony and lower respondent burden could be obtained at little or no cost.

Results

The hypothesized measurement model was tested first with the calibration sample. This random third of the full dataset contained 391 cases. It did not differ significantly from the two validation samples on gender, race/ethnicity, free/reduced lunch program participation, or teacher's report that students were below, at, or above grade level. Rates of missing values were low: on average 1.6 values (2%) out of the analyzed 80 variables were missing per case.

Demographic data collected from caregivers with the ESSP for Families were available for 58.6% of the calibration sample. According to caregivers, 45.4% of the sample members with demographic data were boys and 54.6% were girls. Over a third of the sample was African American (37.1%), 8.3% was Hispanic/Latino, and 47.6% were European American. Just over 40% (41.2%) of the children received free or reduced price lunches at school. School performance data collected from teachers were available for 90.8% of the calibration sample. According to teachers, from whom data were available for 91% (n=354) of the child cases, just under one third of the sample was below grade level in reading (31.6%) and math (31.4%); 20.9% of the children were above grade level in reading and 16.6% were above in math. The rest were at grade level in each subject area.

When the full model with 80 variables representing 12 hypothesized dimensions was run, fit was inadequate according to most of our pre-established criteria (χ^2= 22.52, df = 8, p = .004; CFI = .93, TLI = .93, RMSEA = .068, WRMR = 1.35). In addition, three standardized factor loadings were under .40, 15 squared multiple correlations (item R^2's) were under .33 (meaning less than one third of their variance was explained by the latent factor with which they were associated), and many residual correlations were over .10. Because so many items had one or more undesirable performance indicators, making a systematic approach to identifying and addressing sources of poor fit difficult, we used Bollen's (2000) more incremental approach described above.

Modifications were primarily item deletions. Items were typically deleted for performing poorly according to more than one criterion, but in general, the reasons applied in the following order of frequency: high residuals, low indicator R^2, double loading(s) as indicated in modification indices, and low primary factor loading. Items were deleted when model fit was inadequate and one or more of these conditions applied. Very few items were deleted for primary loadings under .40. In the model containing all factors, secondary loadings that were small (e.g., .20) but statistically significant were detrimental to overall fit. Because loadings under .40 were not desired, ruling out the option of allowing cross-loadings, few options were available except deletion. The use of the polychoric correlation matrix, which generally contains higher correlations than the corresponding Pearson correlation matrix, and the sample size likely contributed to the double-loading problem. A more major modification was the creation of two dimensions (Friends Have Good Behavior and Friends Have Bad Behavior) to replace of the one that was hypothesized. The final model had 61 indicators of 13 dimensions in five domains.

The final good-fitting model obtained in the calibration sample through the jigsaw process was validated with the two validation samples. Table 1 presents the fit statistics for all three

samples. A non-significant *p* value was obtained in each sample. In addition, the CFI and TLI values were high, and RMSEA values were below .05, the most conservative of common recommended cut offs. The WRMR value was above the cutoff recommended at the Mplus website (L. K. Muthén & Muthén, 2010). However, following other recommendations at the Mplus website for situations in which the RMSEA or WRMR demonstrates anomalous scores as discussed above (L. K. Muthén & Muthén, 2010), we accepted the model as a good fitting model. The non-significant χ^2 and consistently excellent fit according to the other three fit measures in three independent samples were compelling reasons for this conclusion.

The factor structure of the final model is presented in Figure 1. Each dimension is summarized in Table 2, and more information on the statistical performance of dimensions in presented in Table 3. The model had 13 dimensions measured with 61 of the original 80 items. Only three of the 13 factors had either one or two loadings below .60. Although these loadings suggest that the indicators are weaker than others in the model, they are statistically significant, and over the typical .40 standard used in exploratory factor analysis.

Three items were removed from the Neighbors Who Care dimension. Retained items, which had loadings from .41 to .69, have to do with adults being supportive to children. Deleted items reflected children's perceptions that adult knew them and were nice to them. Although the construct became more narrow as a result of item deletions, the alpha coefficients obtained from the three random test samples were consistently below the minimum desirable value of .70. As shown in Table 3, the average alpha from three samples was .63.

In the school domain, one item about whether the teacher calls on the student when he/she raises his/her hand was deleted from the Teachers Who Care dimension. It is possible that the item did not factor well because even caring teachers cannot call on all students who raise their hands. Remaining items had loadings over .60. Items on a Fun Place to Learn performed adequately in the CFA, with loadings between .52 and .88. One item on the Fun Place to Be With Other Kids was deleted; scale reliability was higher without the item. It is possible that the expression "look forward to" made this item difficult for some children. Remaining items had loadings of .67 or higher. Average alphas for the three scales in the School domain were .72, .71, and .71 respectively.

In the Friends domain, the dimension Friends Who Care factored well with all loadings .63 or higher, and had an average alpha of .74. Two items were removed from Accepted by Other Children, both of which had to do with being picked on in places other than school. Remaining items had loadings over .70 and the scale's average alpha was .85. Items measuring the hypothesized Friends Have Good Behavior dimension loaded onto two factors, one with four negatively worded items (which was named Friends Have Bad Behavior) and one with four positively worded items (Friends Have Good Behavior). One item, My friends fight with other kids, did not perform well and was removed from the Bad Behavior scale. The two Friend Behavior scales had adequate reliability: .84 (Good Behavior) and .75 (Bad Behavior).

The Family domain of the ESSP for Children contains one dimension, Family Who Care. Two deleted items were related to family members doing fun things together, and playing and laughing together. The remaining items had loadings over .70 and alpha's for the scale from the three test samples averaged .82.

In the Health and Well-being domain, Good Physical Health had five indicators in the final model, all of which refer specifically to feeling sick or having aches (headache, stomachache, toothache). Three other items had low R^2's and high residuals and were removed from the scale. These items (having trouble seeing and hearing in the classroom,

and having trouble getting to sleep at night) appear to assess a broader physical health construct. The items related to seeing and hearing in the classroom are reported on individual ESSP profiles, even though they are not included in the scale. Remaining scale items loading at .60 or higher on the factor, and the average for the scale was .69.

Positive Feelings about Self was reduced by three items that focused on specific skills—in music, art, and sports due to poor performance (low R^2's, high residuals). The remaining more general items (e.g., I am a nice person, a good friend) have loadings of .70 or higher and an average alpha of .75. The third scale in the Health and Well-being domain is Good Adjustment. Five of six items had high loadings (.63 or higher), and no double loadings. One item was removed from the scale due to a low R^2 and high residuals: Do you ever worry bad things will happen.

Knows Where to get Support is the final dimension in the Health and Well-being domain. Two items were removed; one for double loading and the other for high residuals. One of the deleted items was excessively lengthy--Someone helps me out (for example, by giving me a ride, helping me with my homework, or giving me a little money). Remaining items had loadings ranging from .48 to .78 and an average alpha in the three test samples of .73.

Table 3 also presents inter-factor correlations, factor variances, and the range of indicator R^2's (squared multiple correlations) obtained in the calibration sample for each factor. All factor variances were significantly different from 0, indicating that the measures capture significant differences in the latent variables across individual children. Correlations among the factors ranged from .03 and .45, indicating that the measured constructs are independent from each other. Neighbors Who Care and Good Physical Health had a pattern of relatively lower correlations with other factors. Family Who Care, Positive Feelings About Self, and Knows Where to Get Support had general patterns of relatively higher correlations with other constructs.

Discussion

To help schools identify and appropriately address influences that contribute to persistent low academic achievement and performance gaps among subgroups of elementary school students, school professionals need high quality assessments of the social environment and the personal well-being they feel in the environment. The ESSP is unique in its broad coverage of influential social environmental influences. In addition, its inclusion of children as a source of information on their experiences in the domains of neighborhood, school, peer system, family, and well-being increase its potential as an assessment tool for school social workers. The current study focused on the quality of the "internal structure" (American Educational Research Association et al., 1999) of data collected online with the ESSP for Children. Using sophisticated confirmatory factor analytic techniques appropriate for clustered, ordinal data with missing values, the study found that ESSP data from children could be modeled as 13 dimensions in five domains. High standards were used for the evaluation of fit and the model was validated with two independent samples.

The factor structure of the final model had one more factor than the hypothesized model. The hypothesized friend behavior construct factored into two separate constructs—one containing negative behavior indicators and the other containing positive behaviors. This suggests that "bad" behaviors, such as hitting and telling lies, are not just the opposite of positive behaviors, such as helping adults and following rules. Instead, antisocial and prosocial behaviors are qualitatively different and can co-exist in the same child.

The magnitude of obtained factor loadings was adequate for all indicators in the final model, suggesting that indicators represented factors well. In addition, there were no indicators that

loaded on more than one factor or that had correlated errors. The internal structure of the ESSP for Children, therefore, appears robust. Latent factors demonstrated meaningful variance in children's perceptions of their environments and their own well-being. Variance in the factors is important for the practice-related validity (N. K. Bowen, 2008b) of the ESSP for Children because it indicates that the instrument can help school social workers distinguish among children with high and low levels of constructs that are theoretically and empirically significant to developmental outcomes.

Inter-correlations among factors indicated that each factor measured a construct that was distinct from the other measures in its domain and on the child-report instrument as a whole. Although the organization of the ESSP for Children by domains (Neighborhood, School, Friends, Family, and Well-being) is useful in the presentation of the online items and the reporting of results, dimensions within domains were not necessarily more highly correlated than dimensions across domains. Nor were they expected to be. For example, we did not expect that children who enjoy the social aspects of school (Fun Place to be With Other Children) would necessarily also score highly on school as a Fun Place to Learn. Similarly, In the Friends domain, child reports of their friends' positive behaviors was not expected to correlate highly with Accepted by Other Children, which assesses (and reverse-codes) the unkind behaviors of "other kids," as opposed to friends. The variations in correlations within and across dimensions may reflect the ongoing reciprocal and transactional influences and fluctuations in students' experiences with other individuals in their environments. They highlight the importance of assessing multiple dimensions of each environment (across sources, and/or within), and acknowledging that each environment may contain risk as well as promotive factors.

Cronbach alphas estimated in the three test samples ranged from .60 to .85, with data from only one of the 13 scales consistently demonstrating a coefficient well below the pre-established minimal standard of .70. The low alphas for the Neighbors Who Care scale, which come after multiple previous attempts to improve its reliability, suggest that children in the age range targeted by the ESSP are not capable of reporting reliably on the behaviors of neighborhood adults. A previous study found that teachers tend to assume student neighborhoods are more negative than what parents and students report (N. K. Bowen & Powers, 2005). To counter potentially inaccurate assumptions about the neighborhood, school staff using the ESSP may need to rely more heavily on parent-report neighborhood data than child data.

The reliabilities of the scales on the ESSP for Children compare favorably with other child report scales focusing on selected dimensions of well-being and the social environment, as discussed in a previous report (N. K. Bowen, 2006). Additional studies that support the claim include Measelle et al. (2005), who report child alphas on personality traits, and Brock et al. (2008), who report alphas obtained with child data on school environmental measures.

In evaluating factor loadings, indicator R^2's, and internal consistency coefficients in combination, it appears that the constructs that yield the strongest data on the ESSP for Children are Accepted by Other Children, Friends Have Bad Behavior, Family Who Care, and Positive Feelings about Self. The quality of the scores on peer and well-being scales may be especially important because previous research (N. K. Bowen & Powers, 2005) showed that teachers' "knowledge" in these two areas in the absence of social environmental data were least likely (after the neighborhood scale) to match actual child reports. In addition, in the absence of social environmental assessment data, teachers were likely to *over*-estimate the positive characteristics of student experiences in the family and with peers (N. K. Bowen & Powers). In terms of peers, others have also noted that teacher ratings of student experiences with negative peer behaviors may be hindered by the covert nature of

those behaviors (Pepler & Craig, 1998), which makes the child reports all the more important. Data on child perceptions of school as A Fun Place To Learn and A Fun Place to be with Other Children, and their own Good Physical Health may not be as strong as other dimensions, but are adequate, especially in the context of a multiple source assessment.

Strengths of the study include the large, diverse sample, and the use of statistical features for ordinal and clustered data. In addition, high standards were used for assessing model fit. Limitations of the study include the fact the ESSP sample with child data available for analysis differed from the full sample on the percentage of children who were below grade level. The analyzed sample had fewer children who were below grade level, according to teacher report, than the full sample. This fact suggests that children with lower school performance may be less likely to take (or to be asked to take) the ESSP. This phenomenon warrants future study. In addition, lower response rates among parents than among children and teachers reduce our knowledge of demographic characteristics of the full analyzed sample.

The current CFA of the ESSP for Children represents only one of many necessary steps for building evidence that data from the instrument have validity for school practice. Future research should examine group differences in scores, and relationships of scores to outcome data from parents, teachers, and other sources.

Practice Implications

By documenting the psychometric qualities of the ESSP for Children, the current study also supports claims that the data have practice-related validity, i.e., scores from the ESSP can contribute to decision-making about student needs and intervention choices related to school success. Child-report information from the ESSP can help fill the need in schools for data that are "relevant to the outcomes of interest in the setting" (N. K. Bowen & Powers, 2005), because each of the assessed dimensions is associated with school performance in the empirical and theoretical literature. The analyzed sample included sizable percentages of African American and Hispanic/Latino students, and students participating in the school lunch program. Therefore, the current study also suggests that the ESSP for Children collects data that can be used to identify needs related to causes of the achievement gap.

Child-report data from the ESSP may also support efforts to establish assessment-driven, evidence-based practices in schools (Schaughency & Ervin, 2006; Stollar, Poth, & Cohen, 2006). Consistent with current beliefs that the validity of assessment data relates to how data are used to benefit respondents (American Educational Research Association et al., 1999), scores on each dimension of the ESSP for Children can be directly linked to effective practices through a publicly available online database (http://www.schoolsuccessonline.com/), which has been described elsewhere (Powers, Bowen, & Bowen, in press). Other online databases of effective practices also have features that facilitate the identification of interventions targeting specific risk factors that may be identified with the ESSP (e.g., http://www.colorado.edu/cspv/blueprints/; http://www.promisingpractices.net/programs.asp).

In sum, according to the current study, the child-report component of the ESSP provides social environmental data of good quality, with which school social workers can identify leverage points for improving the school performance of 3rd through 5th graders from diverse backgrounds. When used as a starting point in an assessment-driven, evidence-based process, the data can benefit students by leading to appropriate and effective strategies to address threats to their school success.

References

Achenbach TM, McConaughy SH, Howell CT. Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. Psychological Bulletin. 1987; 101:213–232. [PubMed: 3562706]

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. Standards for educational and psychological testing. Washington DC: American Educational Research Association; 1999.

Baydar N, Brooks-Gunn J, Furstenberg F. Early warning signs of functional illiteracy: Predictors in childhood and adolescence. Child Development. 1993; 64:815–829. [PubMed: 8339697]

Bollen, KA. Structural equations with latent variables. New York: John Wiley & Sons; 1989.

Bollen KA. Modeling strategies: In search of the Holy Grail. Structural Equation Modeling. 2000; 7:74–81.

Bowen, GL.; Richman, JM.; Bowen, NK. The School Success Profile: A results management approach to assessment and intervention planning. In: Roberts, AR.; Greene, GJ., editors. Social workers' desk reference. New York: Oxford University Press; 2002. p. 787-793.

Bowen, GL.; Rose, RA.; Bowen, NK. The reliability and validity of the School Success Profile. Philadelphia, PA: Xlibris; 2005.

Bowen NK. Psychometric properties of the Elementary School Success Profile for Children [Instrument Development]. Social Work Research. 2006; 30:51–63.

Bowen NK. Cognitive testing and the validity of child-report data from the Elementary School Success Profile. Social Work Research. 2008a; 32:18–28. [PubMed: 20148125]

Bowen, NK. Validation. In: Darity, WA., Jr, editor. International encyclopedia of the social sciences. 2. Vol. 8. Detroit: Macmillan Reference; 2008b. p. 569-572.

Bowen, NK.; Bowen, GL.; Woolley, ME. Constructing and validating assessment tools for school-based practitioners: The Elementary School Success Profile. In: Roberts, AR.; Yeager, KR., editors. Evidence-based practice manual: Research and outcome measures in health and human services. New York: Oxford University Press; 2004. p. 509-517.

Bowen NK, Powers JD. Knowledge gaps among school staff and the role of high quality ecological assessments in schools. Research on Social Work Practice. 2005; 15:491–500.

Brock LL, Nishida TK, Chiong C, Grimm KJ, Rimm-Kaufman SE. Children's perceptions of the classroom environment and social and academic performance: A longitudinal analysis of the contribution of the Responsive Classroom approach. Journal of School Psychology. 2008; 46:129–149. [PubMed: 19083354]

Corcoran, K.; Fischer, J. Measures for clinical practice: A sourcebook. 3. Vol. 1. New York: Free Press; 2000.

DeVellis, RF. Scale development: Theory and applications. 2. Thousand Oaks, CA: Sage; 2003.

Flora DB, Curran PJ. An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods. 2004; 9:466–491. [PubMed: 15598100]

Franklin, C.; Corcoran, J.; Harris, MB. Risk and protective factors for adolescent pregnancy: Bases for effective intervention. In: Fraser, MW., editor. Risk and resilience in childhood: An ecological perspective. Washington D.C: NASW Press; 2004. p. 281-313.

Holmbeck, GN.; Greenley, RN.; Franks, EA. Developmental issues in evidence-based practice. In: Barrett, PM.; Ollendick, TH., editors. Handbook of interventions that work with children and adolescents: Prevention and treatment. Chichester, West Sussex, England: John Wiley; 2004. p. 27-48.

Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling. 1999; 6:1–55.

Jenson, JM. Risk and protective factors for alcohol and other drug use. In: Fraser, MW., editor. Risk and resilience in childhood: An ecological perspective. 2. Washington, DC: NASW Press; 2004. p. 183-208.

Jöreskog, KG. Testing structural equation models. In: Bollen, K.; Long, JS., editors. Testing structural equation models. Newbury Park: Sage Publications; 1993. p. 294-316.

Jöreskog, KG. Structural equation modeling with ordinal variables using LISREL. 2005. Retrieved March 9, 2007, from www.ssicentral.com/lisrel/techdocs/ordinal.pdf

Kline, RB. Principles and practice of structural equation modeling. 2. New York: Guilford; 2005.

Lloyd DN. Prediction of school failure from third-grade data. Educational and Psychological Measurement. 1978; 38:1193–1200.

Measelle JR, John OP, Ablow JC, Cowan PA, Cowan CP. Can children provide coherent, stable, and valid self-reports on the big five dimensions? A longitudinal study from ages 5 to 7. Journal of Personality and Social Psychology. 2005; 89:90–106. [PubMed: 16060748]

Merrell, KW. Behavioral, social, and emotional assessment of children and adolescents. 2. Mahwah, NJ: Lawrence Erlbaum; 2003.

Muthén, LK.; Muthén, BO. Mplus user's guide. 5. Los Angeles, CA: Muthén & Muthén; 1998–2007.

Muthén, LK.; Muthén, BO. Mplus (Version 4.2). Los Angeles, CA: Muthén & Muthen; 2004.

Muthén, LK.; Muthén, BO. 2010. http://statmodel.com/

National Assessment of Educational Progress. The Nation's Report Card: Grade 4. Mathematics: U.S. Department of Education; 2009a.

National Assessment of Educational Progress. The Nation's Report Card: Grade 4, Reading: US Department of Education. 2009b.

Offord DR, Boyle MH, Racine Y, Szatmari P, Fleming JE, Sanford M, et al. Integrating assessment data from multiple informants. Journal of the American Academy of Child and Adolescent Psychiatry. 1996; 35:1078–1085. [PubMed: 8755805]

Pepler DJ, Craig WM. Assessing children's peer relationships. Child Psychology & Psychiatry Review. 1998; 3:176–182.

Powers, JD. Evidence-based practice in schools: Current status, potential barriers, and critical next steps. University of North Carolina at Chapel Hill; Chapel Hill, NC: 2005.

Powers JD, Bowen NK, Bowen GL. Supporting evidence-based practice in schools with an online database of best practices. Children & Schools. in press.

Reynolds, CR.; Kamphaus, RW. Handbook of psychological and educational assessment of children: Personality, behavior, and context. 2. New York: Guilford; 2003.

Richman, JM.; Bowen, GL.; Woolley, ME. School failure: An eco-interactional developmental perspective. In: Fraser, MW., editor. Risk and resilience in childhood: An ecological perspective. 2. Washington DC: NASW Press; 2004. p. 133-160.

Sameroff, AJ. Ecological perspectives on developmental risk. In: Osofsky, JD.; Fitzgerald, HE., editors. WAIMH handbook of infant mental health: Infant mental health in groups at high risk. Vol. 4. New York: John Wiley & Sons; 2000. p. 1-33.

Schaughency E, Ervin R. Building capacity to implement and sustain effective practices to better serve children. School Psychology Review. 2006; 35:155–166.

Silverman, WK.; Saavedra, LM. Assessment and diagnosis in evidence-based practice. In: Barrett, PM.; Ollendick, TH., editors. Handbook of interventions that work with children and adolescents: Prevention and treatment. Chichester, West Sussex, England: John Wiley; 2004. p. 50-69.

SPSS Inc. SPSS for Windows (Version 14.0). Chicago: SPSS Inc; 2005.

Stollar SA, Poth RL, Curtis MJ, Cohen RM. Collaborative strategic planning as illustration of the principles of systems change. School Psychology Review. 2006; 35:181–197.

U.S. Census Bureau. Current Population Survey, 2009 Annual Social and Economic Supplement. Washington, DC: author; 2008. Table PINC-03. Educational Attainment--People 25 Years Old and Over, by Total Money Earnings in 2008, Work Experience in 2008, Age, Race, Hispanic Origin, and Sex.

U.S. Department of Agriculture. Child Nutrition Programs—Income Eligibility Guidelines. Vol. 72. Washington, DC: Federal Register; 2007. p. 8685

Wegmann KM, Thompson AM, Bowen NK. Confirmatory factor analysis of home environment and home social behavior data from the ESSP for Families. Social Work Research. in press.

Williams, JH.; Ayers, CD.; Van Dorn, RA.; Arthur, MW. Risk and protective factors in the development of delinquency and conduct disorder. In: Fraser, MW., editor. Risk and resilience in childhood: An ecological perspective. 2. Washington D.C: NASW Press; 2004. p. 209-249.

Woolley ME, Bowen GL, Bowen NK. Cognitive pretesting and the developmental validity of child self-report instruments: Theory and applications. Research on Social Work Practice. 2004; 14:191–200.
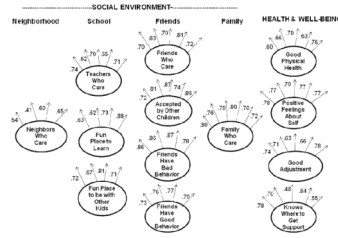
Figure 1.
Final CFA Model of the ESSP for Children with Standardized Factor Loadings

Table 1

Fit Statistics for Final Model in Three Independent Samples

Fit statistics and *desired values*	χ^2 (df) (p-value) *non-significant*	CFI *>.95*	TLI *>.95*	RMSEA *<.05 – .08*	WRMR *<.90*
Calibration Sample	15.032 (8) (0.059)	0.973	0.973	0.047	1.030
Validation Sample 1	16.447 (9) (0.058)	0.969	0.969	0.046	1.021
Validation Sample 2	14.833 (9) (0.096)	0.979	0.977	0.041	0.976

Note: The number of degrees of freedom in Mplus when the WLSMV estimator is used is different from the usual difference in non-redundant covariance matrix elements and estimated parameters (L.K. Muthén & Muthén, 1998–2007). The number of free parameters in all models was 312.

Table 2

Summary of 61 Items in 13 Dimensions In Final CFA Model

Neighbors Who Care: 4 items assess the child's view that adults in the neighborhood care about the child, are responsive to the child, and watch out for the child.

Teachers Who Care: 5 items assess the child's view that the teacher is responsive to him or her in the classroom and cares about him or her.

A Fun Place to Learn: 4 items assess whether the child looks forward to going to school and learning new things.

A Fun Place to be with Other Children: 4 items assess whether the child has friends to talk to, play with, and go to lunch with at school, and whether the child has fun with friends at school.

Friends Who Care: 5 questions assess whether the child believes his or her friends are emotionally supportive and fun to play with.

Accepted By Other Children: 5 questions assess the child's view of how he or she is treated by other children, e.g., do other kids tease, make fun of, or exclude child from play.

Friends Have Bad Behavior: 4 items assess the child's view that friends engage in negative behaviors, such as hitting others, hurting other kids' feelings, and telling lies

Friends Have Good Behavior: 4 questions assess the child's view that friends engage in positive behaviors, such as helping adults and following the rules.

Family Who Care: 6 items assess the child's view that his or her family is emotionally supportive, and that adults at home make him or her feel special and are responsive when child is upset.

Good Physical Health: 5 items assess the child's view of his or her health (e.g., having toothaches, headaches, or stomachaches, feeling sick or sleepy).

Positive Feelings About Self: 5 items assess the child's view of his or her competence in different domains—being smart, a good friend, a nice person.

Good Adjustment: 5 items on assess the child's feelings of being confused, uncared for, alone, or scared.

Knows Where to Get Support: 5 items assess the child's view that people are available to talk to and provide praise and moral support.

Table 3

Additional Factor Model Statistics: Inter-factor Correlations,[*] Factor Variances,[^] and Explained Variance[#] of Observed Indicators

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Neighbors Who Care													
2. Teachers Who Care	.17												
3. Fun Place to Learn	.11	.25											
4. Fun Place/Other Children	.11	.25	.13										
5. Friends Who Care	.16	.29	.17	.40									
6. Accepted by Other Children	.03	.18	.04	.23	.14								
7. Friends/Bad Behavior	.06	.16	.03	.16	.27	.24							
8. Friends /Good Behavior	.13	.28	.21	.17	.31	.04	.36						
9. Family Who Care	.18	.42	.26	.26	.35	.16	.25	.36					
10. Good Physical Health	.03	.05	.13	.09	.12	.20	.24	.10	.12				
11. Positive Feelings/Self	.19	.39	.28	.34	.39	.10	.27	.41	.45	.11			
12. Good Adjustment	.06	.23	.12	.16	.20	.29	.24	.16	.34	.23	.20		
13. Knows Where/Support	.21	.39	.22	.38	.40	.15	.21	.31	.53	.12	.44	.25	
Factor Variances	.29	.54	.39	.52	.50	.52	.74	.53	.64	.36	.60	.55	.60
Average alphas (3 samples)	.63	.72	.71	.71	.74	.85	.84	.75	.82	.69	.75	.75	.73
Range of Indicator R²'s	.17–.48	.39–.54	.27–.78	.45–.65	.39–.66	.52–.75	.61–.75	.53–.62	.49–.64	.36–.57	.50–.60	.39–.61	.23–.60

[*] All correlations higher than .07 were statistically significant (critical ratios > 1.96).

[^] All factor variances were statistically significantly different from 0 (critical ratios > 1.96).

[#] The reported indicator R²'s are from the calibration sample.